

J_073

主成分分析を用いた動画像中の字幕の除去
Elimination of Superimposition in Videos Using Principal Component Analysis

尾関 孝史† 福永 裕也† 小林 富士男†
Takashi Ozeki Yuya Fukunaga Fujio Kobayashi

1. まえがき

テレビ画像や映画には、しばしばオリジナル映像に字幕などが合成されている。この字幕が含まれた映像から字幕を除去し、オリジナルの映像を復元することを試みる。本発表では、動画像の連続するフレームから、学習サンプルを作成し、主成分分析を用いて固有部分空間を作成する。その固有部分空間に字幕により欠損した部分画像を逆射影することで映像の復元を行う。提案手法は静止画像から文字等を除去する BPLP 法[1]の拡張であるが、BPLP 法で仮定する画像の自己相関性やフラクタル性を仮定しない。さまざまな動画に字幕を付け加え、提案手法でどの程度復元が可能であるかを実験した結果を報告する。

2. 提案手法

BPLP 法は、主成分分析を用いて、1枚の静止画像からテロップを除去しようとする方法である。この方法は、画像に自己相関性やフラクタル性がある場合は、非常に良好にテロップを除去できることが知られている[1]。1枚の静止画像では、自己相関性が低い場合でも、動画では、しばしば連続するフレーム間の相関性は高い。そこで、BPLP 法を拡張し、前後のフレームの情報を利用して、動画像に含まれる字幕の除去を試みる。

BPLP 法では、復元したいテロップを含む部分の情報はいらず、テロップを含まない他の部分領域から画像が持つ特徴部分空間を抽出している。そして、テロップによる欠損画素を含む部分領域を特徴部分空間へ逆写像することでテロップの除去を行っている。これに対して、提案手法では、欠損画素の情報は、前後のフレームの欠損画素を含む部分領域に最もあると考え、この部分領域から特徴部分空間を作成することにする。

動画像の画像サイズを $W \times H$ とする。復元しようとする欠損領域を含むフレームを F_i とする。また、図1に示すように、画像サイズより小さなウィンドウ $w \times h$ ($w < W, h < H$) で画像を分割し、字幕による欠損領域を含む部分領域をラスタスキャンしたベクトルを \hat{x} とする。

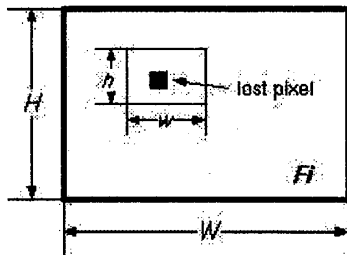


図1 欠損領域のベクトル \hat{x} の切り出し
このとき、カラー画像では、RGBRGB...と輝度値を RGB

の順に並べてサンプルベクトルを作成する。

次に、図2に示すようにフレーム F_i の前後のフレームから、 \hat{x} と同じ位置のウィンドウおよびそれを前後左右斜めに数画素分、平行移動したウィンドウからラスタスキャンして作成した学習サンプルを

$$x_i = [x_{i1}, x_{i2}, \dots, x_{i_{w \cdot h}}]^T \quad (i=1, 2, \dots, N)$$

とする。このようにして得られた学習サンプルから画像ベクトル列

$$X = [x_1, x_2, \dots, x_N]$$

を作成する。なお、各学習サンプル x_i の欠損画素に対応する成分の値に関しては、その同じ成分が欠損画素でない、他の学習サンプルベクトルの成分の平均値とする。

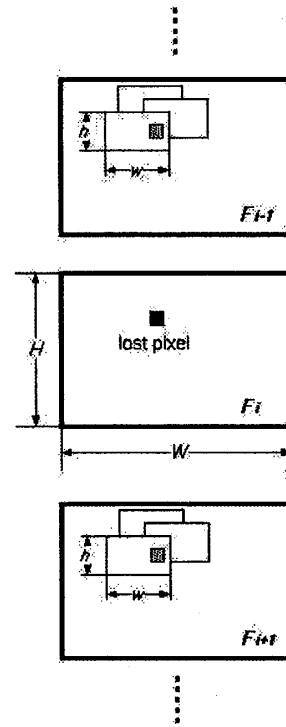


図2 学習サンプルの作成

これ以後は、BPLP 法と同様にして画像復元を行う。共分散行列

$$\frac{1}{N-1} XX^T$$

に対して主成分分析を行い、得られた固有ベクトル列を

$$E = [e_1, e_2, \dots, e_D]$$

とする。なお、実際の計算では、行列 X の特異値分解を行い、左特異値ベクトルを求めることで、共分散行列の固有ベクトルを求める。また、 D は固有部分空間の次元で、

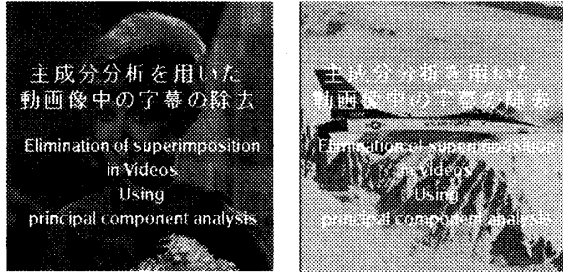
† 福山大学工学部: Faculty of Engineering, Fukuyama University

特異値の累積寄与率を参考に決定する。そして、欠損領域を復元したウィンドウ x^* を

$$x^* = E(E^T \Sigma E)^{-1} E^T \hat{x}$$

で得る。ここで、 Σ は欠損画素の有無を示す $w \cdot h$ の大きさの対角行列で、 \hat{x} の欠損画素に対応する対角成分は0、そうでない成分は1を取る。

3. 実験結果



(a) Girl (b) Airplane
図3 字幕のある画像

実験では大きさが 256×256 , 256 階調のカラー標準画像の Girl と Airplane に対して、図3のような字幕を合成した。字幕による欠損画素の総数は 4413 である。実験のための動画として、

- ① 字幕、背景ともに動かない静止画像
 - ② 字幕のみ上から下に1画素ずつ移動し、背景は動かない動画像
 - ③ 背景のみ下から上に1画素ずつ移動し、字幕は動かない動画像
- の3種類を 11 フレーム用意した。ここで、3つのどの動画においても中央のフレームの画像は同じになるようにした。

また、学習サンプルに関しては、字幕を含む復元しようとするウィンドウと同じ位置の他のフレームのウィンドウから作成した学習サンプルとそのウィンドウを上下左右斜めに1又は2画素分、平行移動した 25 個のウィンドウを 10 フレームから作成した合計 250 個を用いた。ここで、ウィンドウの大きさは 16×16 画素とし、対象のフレームを 16×16 個のウィンドウに分割し、欠損画素が含まれているウィンドウに対しての復元をそれぞれ行った。また、固有部分空間の次元は、Girl, Airplane 共に累積寄与率が 60% までの次元を採用した。そして、BPLP 法に関しては、中央のフレームのみから字幕の除去を試みた。BPLP 法の固有部分空間の次元も、提案手法と同じように累積寄与率が 60% までの次元を採用した。

また、復元画像の評価は平均二乗誤差 MSE

$$MSE = \frac{1}{3N(\Omega)} \sum_{(m,n) \in \Omega} |x(m,n) - x^*(m,n)|^2$$

を用いた。ここで、 $x(m,n)$ は字幕を合成する前の標準画像の輝度値であり、 $x^*(m,n)$ は復元画像の輝度値である。また、集合 Ω は、欠損画素の集合で、 $N(\Omega)$ は字幕による欠損画素の総数である。以下、3種類の動画に対する実験結果を述べる。

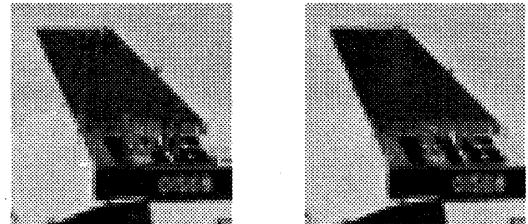
表1によると、Girl, Airplane の両方とも、3種類すべての動画像に対して、提案手法による復元画像のほうが、

静止画1枚から復元を行う BPLP 法よりも平均二乗誤差の値が小さく、字幕の除去が良好な結果となった。特に、動画像①に関しては、すべてのフレームが同じ画像の静止画であり、BPLP 法で使用される画像の情報と動画像の持つ画像の情報の間に差はない。BPLP 法は学習サンプルを欠損画素のない離れた領域のウィンドウから作成するので、復元しようとするウィンドウと学習サンプルとの間に強い相関性が必要となる。一方、提案手法では、学習サンプルは欠損画素を含むウィンドウから作られ、欠損画素部分は、そのウィンドウを上下左右斜めに数画素分移動したウィンドウの同成分の値の平均を取るようになっている。このことは、提案手法では、静止画像の場合、欠損画素の値として、欠損画素の周辺の値のある種の平均を取っていることになる。

表1 原画像と復元画像の平均二乗誤差 MSE

	Girl	Airplane
BPLP Method	105.5	495.6
Proposed Method for ①	78.5	306.7
Proposed Method for ②	34.1	156.4
Proposed Method for ③	64.1	261.1

図4は、BPLP 法と提案手法により、動画像①の Airplane の字幕除去を行った結果の拡大図である。提案手法では、尾翼の部分の復元もうまくできている。



(a) BPLP 法 (b) 提案手法
図4 字幕を除去した画像の拡大図

4. むすび

本発表では、1枚の画像からテロップなどによる欠損画素を修復する BPLP 法を拡張し、動画像の字幕を除去する手法を提案した。動画像の連続するフレームは、相関性が高いため、BPLP 法と異なり、画像自身には自己相関性が高い必要はない。そこで、提案手法では前後のフレームの欠損画素を含むウィンドウから、学習サンプルを作成し、字幕を除去した。静止画像、字幕のみ移動する動画像、背景のみ移動する動画像の3種類の動画像を作成し、実験を行ったところ、前後のフレームの情報を利用する提案手法のほうが BPLP 法より良好な字幕の除去ができることが示された。

参考文献

[1] 天野敏之, 佐藤幸男, “固有空間法を用いた BPLP による画像補間,” 信学論(D-II), Vol. J85-D-II, No. 3, pp. 457-465, 2002.